

## Review

# How do thermophilic proteins deal with heat? \*\*

S. Kumar<sup>a</sup> and R. Nussinov<sup>b,c,\*</sup>

<sup>a</sup> Laboratory of Experimental and Computational Biology, NCI-Frederick, Bldg 469, Rm 151, Frederick, Maryland 21702 (USA)

<sup>b</sup> Intramural Research Support Program-SAIC, National Cancer Institute-Frederick, Frederick Cancer Research and Development Center, Bldg 469, Rm 151, Frederick, Maryland 21702 (USA), Fax + 1 301 846 5598, e-mail: ruthn@ncifcrf.gov

<sup>c</sup> Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978 (Israel)

Received 28 February 2001; received after revision 26 March 2001; accepted 27 March 2001

**Abstract.** Recent years have witnessed an explosion of sequence and structural information for proteins from hyperthermophilic and thermophilic organisms. Complete genome sequences are available for many hyperthermophilic archaeons. Here, we review some recent studies on protein thermostability along with work from our laboratory. A large number of sequence and structural factors are thought to contribute toward higher intrinsic thermal stability of proteins from these organisms. The most consistent are surface loop deletion, increased occurrence of hydrophobic residues with branched side chains and an increased proportion of charged residues at the expense

of uncharged polar residues. The energetic contribution of electrostatic interactions such as salt bridges and their networks toward protein stability can be stabilizing or destabilizing. For hyperthermophilic proteins, the contribution is mostly stabilizing. Macroscopically, improvement in electrostatic interactions and strengthening of hydrophobic cores by branched apolar residues increase the enthalpy change between the folded and unfolded states of a thermophilic protein. At the same time, surface loop deletion contributes to decreased conformational entropy and decreased heat capacity change between the folded and unfolded states of the protein.

**Key words.** Thermostability; salt bridges; electrostatics; adaptation; protein; temperature; structure.

## Introduction

To survive, living organisms must be able to adapt to their natural environment. Nowhere on earth is this simple evolutionary principle more tested than in high-temperature water containing terrestrial, subterranean, and submarine environments. About 75 species of hyperthermophilic archaea and bacteria have been described so far [1]. These

organisms are found in deep sea vents, submarine hydrothermal areas, continental solfataras, and geothermal power plants. Hence, the question arises as to how these organisms are able to not only tolerate high temperatures but also exploit them to their advantage. Understanding higher-temperature resistance of thermophilic and hyperthermophilic proteins is essential for studies of protein folding and stability, and is critical for designing efficient enzymes that can work at high temperatures [2–5]. The first high-resolution crystal structure (of thermolysin [6]) was published about 25 years ago, when Perutz also commented on the stereochemical basis of the thermostability of ferredoxins and hemoglobin A2 [7]. Since then, several

\* Corresponding author.

\*\* Copyright disclaimer: The publisher or recipient acknowledges right of the U.S. Government to retain a nonexclusive, royalty-free license in and to any copyright covering the article.

studies have focused on the adaptation of proteins to elevated temperatures [8–22]. The stability of thermophilic proteins [15, 21, 22] was attributed to greater hydrophobicity, better atom packing, deletion or shortening of loops [12], smaller and fewer cavities, an increased surface area buried upon oligomerization [23, 24], residue substitution within and outside the secondary structures [9, 12], increased occurrence of proline residues in loops [25–27], decreased occurrence of thermolabile residues [12, 21], increased helical content, increased polar surface area, better hydrogen bonding, and better salt bridge formation [15, 17, 21, 22].

Here, we do not present a comprehensive review of the vast body of literature on thermostability. Instead, we focus on the following. First, we define protein stability in terms of *macroscopic* thermodynamic properties. We describe and compare stability curves of hyperthermophilic, thermophilic, and mesophilic proteins. The second part of this review relates to microscopic sequence and structural differences derived from known high-resolution structures of thermophilic and hyperthermophilic proteins. An increase in the proportion of charged residues and improved electrostatic interactions are among the most consistent mechanisms for increasing protein thermal stability. In addition, several studies have reported an increased occurrence of hydrophobic residues with branched side chains in hyperthermophilic and thermophilic proteins. In the third part, we review studies involving genomes of hyperthermophilic organisms. Although such data are still relatively sparse, initial trends indicate that proteins from hyperthermophilic organisms are shorter and have increased proportions of charged residues. Given the consistent evidence implicating charged residues and electrostatic interactions in protein thermostability, the fourth part addresses the energetic contribution of electrostatics to the stability of hyperthermophilic proteins. Finally, we summarize consistent factors enhancing protein thermostability and discuss how they may work at the microscopic level.

### Macroscopic description of protein stability

The thermodynamic stability of a protein is measured by the Gibbs free energy change between the folded native (N) and unfolded denatured (D) states in the ( $N \rightleftharpoons D$ ) folding reaction. This is also called the Gibbs free energy change for protein unfolding:

$$\Delta G = \Delta H - T\Delta S \quad (1)$$

where  $\Delta H$  is the change in enthalpy and  $\Delta S$  is the change in the entropy between the folded and unfolded states of the protein. The enthalpy and entropy changes between

the folded and unfolded states of the protein are a function of temperature

$$\partial\Delta H/\partial T = \Delta C_p \quad (2)$$

$$\partial\Delta S/\partial T = \Delta C_p/T \quad (3)$$

where  $\Delta C_p$  is the change in the heat capacity of the protein between the folded and unfolded states. Usually, it is assumed to be constant in the temperature range relevant to protein stability studies. For the sake of simplicity, below we refer to the change in enthalpy ( $\Delta H$ ), in entropy ( $\Delta S$ ), and in heat capacity ( $\Delta C_p$ ) between the folded and unfolded states of a protein as the enthalpy change, the entropy change, and the heat capacity change, respectively.

Integrating Eqs 2 and 3, we get the enthalpy and entropy changes at a given temperature ( $T$ ),

$$\Delta H(T) = \Delta H_G - (T_G - T) \Delta C_p \quad (4)$$

$$\Delta S(T) = \Delta H_G/T_G - \Delta C_p \ln(T_G/T) \quad (5)$$

where  $T_G$  is the melting temperature of the protein [ $\Delta G(T_G) = 0$ ] and  $\Delta H_G$  is the enthalpy change at  $T_G$  and  $\ln$  is the natural logarithm.

Let us consider a protein or a protein domain that follows a simple two-state folding process ( $N \rightleftharpoons D$ ). Let us assume that this protein is stable over a certain temperature range and has a constant (greater than zero) heat capacity change in this range. For such a protein, the Gibbs-Helmholtz equation can be used to plot its stability curve [28, 29]. This equation is given by the following formula:

$$\Delta G(T) = \Delta H_G (1 - T/T_G) - \Delta C_p [(T_G - T) + T \ln(T/T_G)] \quad (6)$$

Using Eqs 1, 4, and 5, the Gibbs-Helmholtz equation (Eq. 6) can be easily derived. A typical protein stability curve is shown in figure 1. It presents the variation of the Gibbs free energy change [ $\Delta G(T)$ ] between the native and denatured states of a protein (along the Y-axis) with temperature (along the X-axis). The shape of the stability curve is a skewed parabola and hence it intersects the abscissa at two transition temperature values. The low temperature value is the cold denaturation temperature and the high temperature value is the melting temperature of the protein. In this review, we restrict ourselves to heat denaturation. Three thermodynamic parameters are needed to plot a protein stability curve, the melting temperature ( $T_G$ ), the enthalpy change at  $T_G$  ( $\Delta H_G$ ) and the heat capacity change ( $\Delta C_p$ ). These parameters can be determined by experimental means. The two most common ways of experimentally studying protein stability are via thermal and chemical denaturation. Spectroscopic (CD, fluorescence) and differential scanning calorimetry (DSC) techniques are often used. In studies performed using chemical denaturants such as urea and GdnHCl, the Gibbs free energy change for unfolding of a protein at a given tem-

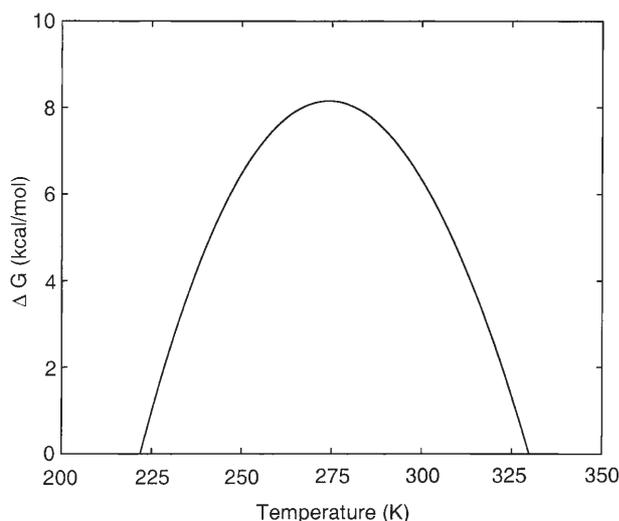


Figure 1. A typical protein stability curve plotted using the Gibbs-Helmholtz equation. This curve represents the variation of Gibbs free energy change  $\Delta G$  between native and denatured states of a reversible two-state folding protein with temperature. Comparison of stability curves for homologous thermophilic and mesophilic proteins may provide useful insights into protein thermostability.

perature (usually room temperature) is estimated most frequently using the linear extrapolation method [30]. This yields two thermodynamic quantities, the heat capacity change ( $\Delta C_p$ ) and the  $m$ -value.  $\Delta C_p$  is the slope of the least-squares line obtained by plotting  $\Delta H$  as a function of  $T$ . It is usually computed around the melting temperature,  $T_G$ . The  $m$ -value is the slope of the least-squares line obtained by plotting the Gibbs free energy change ( $\Delta G$ ) as a function of the denaturant concentration. It is measured around the transition state [30]. In these experimental studies, the crucial parameter to be determined is  $\Delta C_p$ . Both  $\Delta H$  and  $\Delta S$  derive from the heat capacity change [31]. Pace et al. [32] have illustrated that the  $T_G$  values are accurate to better than  $\pm 1\%$  and the value for  $\Delta H_G$  from a van't Hoff analysis can be determined to about  $\pm 5\%$  accuracy, with good agreement between different laboratories. However, there may be considerable differences in  $\Delta C_p$  values.

Analysis of a protein stability curve [28] yields the following quantities that may be useful for comparison among the curves for different proteins.

The slope of the protein stability curve at  $T_G$  is given by

$$\partial \Delta G(T) / \partial T = -\Delta H_G / T_G \quad (7)$$

The curvature of the protein stability curve at  $T_G$  is given by

$$\partial^2 \Delta G(T) / \partial T^2 = -\Delta C_p / T_G \quad (8)$$

Examination of a typical protein stability curve indicates a temperature at which the protein is maximally stable ( $T_S$ ).  $T_S$  is also the temperature at which the entropy change is zero.  $T_S$  is given by

$$T_S = T_G \exp[-\Delta H_G / (T_G \Delta C_p)] \quad (9)$$

At  $T_S$ , the free energy change  $\Delta G(T_S)$  is given by

$$\Delta G(T_S) = \Delta H(T_S) = \Delta H_G - (T_G - T_S) \Delta C_p \quad (10)$$

These macroscopic descriptors of stability are directly related to microscopic structural and sequence properties of the proteins. The heat capacity change of a protein is related to the change in its accessible surface area between the folded and unfolded states and, hence, its sequence length [33]. Recently, Edgcomb and Murphy [34] have reviewed progress in the parameterization of enthalpy, entropy, and heat capacity changes in proteins in terms of their known three-dimensional structures. Ganesh et al. [35] have used a data set of 28 proteins, for which amino acid sequences, X-ray, structures, as well as thermodynamic parameters are known, to analyze the relationship between microscopic and macroscopic properties of proteins. They have shown that it is possible to predict the temperature of maximal stability and molar heat capacity change per residue for the proteins from their amino acid sequences or from their X-ray structures.

Comparison of protein stability curves among hyperthermophilic, thermophilic, and mesophilic proteins provides useful insights into protein thermostability. Beadle et al. [36] have discussed stability curves of thermophile/mesophile pairs. They note that a higher melting temperature in a thermophilic protein can be attained in one of three ways. First, greater maximal stability,  $\Delta G(T_S)$ , of a thermophilic protein can up-shift the stability curve, resulting in a higher melting temperature ( $T_G$ ). Second, thermophilic and mesophilic proteins have similar maximal stabilities. However, the curvatures of the protein stability curves may differ. If the curvature of the stability curve of the thermophilic protein (specified largely by  $\Delta C_p$ ) is smaller, the curve would be broader, leading to a higher melting temperature. Third, there may be a left/right shift of one curve with respect to the other.

Recently, we compared protein stability curves in five homologous families containing hyperthermophilic, thermophilic, and mesophilic proteins. The hyperthermophilic and thermophilic proteins have greater maximal stabilities than their mesophilic homologs. As a result, the protein stability curves of hyperthermophiles and thermophiles are up-shifted and broadened. However, the protein stability curve of a cold-shock protein from *Thermotoga maritima* (CspTm) is both up- and right-shifted. A separate set of two homologous families containing only the mesophilic proteins did not show these trends [Kumar et al., unpublished data].

Our analysis comprised of a total of 30 proteins and several correlations among the various thermodynamic parameters were observed. Among the families containing homologous thermophilic and mesophilic proteins, we observed a correlation between melting temperature ( $T_G$ ) and maximal protein stability [ $\Delta G(T_S)$ ]. Maximal protein stability is in turn correlated with the enthalpy change at

the melting temperature ( $\Delta H_G$ ). The enthalpy change per residue ( $\Delta h_G = \Delta H_G/N_{res}$ ) correlates well with the melting temperature ( $T_G$ ). These correlations yield a consistent picture. Hyperthermophilic proteins gain their greater temperature resistance by increasing their maximal stabilities. Higher maximal stability is achieved by a greater enthalpy change at the melting temperature. The greater enthalpy change is derived from greater (e.g., electrostatic) interactions formed in hyperthermophilic proteins but absent in their mesophilic homologs [Kumar et al., unpublished data].

Several studies have correlated macroscopic differences in stabilities among hyperthermophiles, thermophiles, and mesophiles with the microscopic structural and sequence differences. Here, we summarize three particularly interesting examples. These are the recent studies on ribonuclease H from *Thermus thermophilus* and *Escherichia coli*, and hyperthermophilic and mesophilic archaeal histones and cold-shock proteins from *T. maritima* and *Bacillus subtilis*.

Hollien and Marqusee [37, 38] studied both thermodynamic and structural properties of cysteine-free mutants of *E. coli* and *T. thermophilus* ribonuclease H. The two enzymes share 52% sequence identity and have very similar three dimensional structures. Protein stability curves of the two enzymes are plotted in figure 2. The stability curve of *T. thermophilus* ribonuclease H is up-shifted and broader than that of *E. coli* ribonuclease H indicating its greater stability. Native-state hydrogen exchange experiments were used to identify the residues that contribute toward the stability of *T. thermophilus* ribonuclease

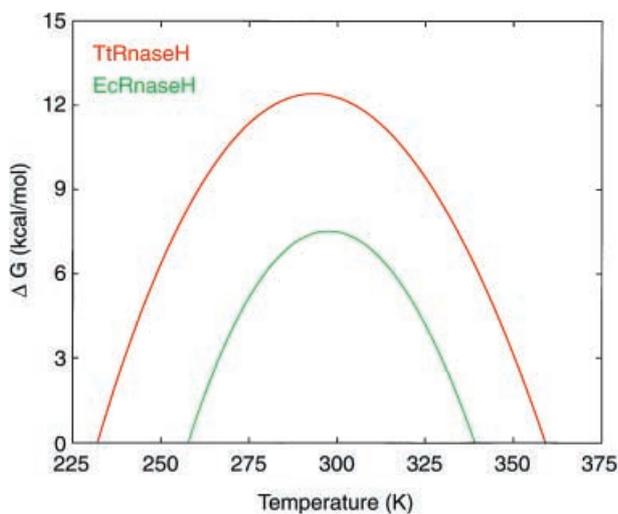


Figure 2. Comparison of protein stability curves for *E. coli* and *T. thermophilus* ribonuclease H shown in green and red, respectively. The stability curve for the thermophilic protein is up-shifted and broadened in comparison to the curve for the mesophilic protein. Data are taken from Hollien and Marqusee [38]. This appears to be a general feature of such comparisons [Kumar et al., unpublished data].

H. These experiments indicate that the greater stability of the thermophilic enzyme is distributed over the protein in a delocalized manner. The protein appears to be stabilized by several local interactions distributed throughout the protein to maintain a stability balance.

Li et al. [39] have studied the thermodynamic stability of related mesophilic and thermophilic archaeal histones. The stabilities of four archaeal histones, rHFoB (from the mesophile *Methanobacterium formicium*, growth temperature 43 °C), rHMfA and rHMfB (from the hyperthermophile *Methanothermus fervidus*, growth temperature 83 °C), and rHPyA1 (from the hyperthermophile *Pyrococcus* strain GB-a, growth temperature 95 °C), were characterized in terms of temperature, salt, and pH. These histones are dimeric and each monomer contains 66–69 residues. The archaeal histones show high sequence and structural conservation. The protein stability curves of the hyperthermophilic histones are up-shifted and broadened compared to that of the mesophilic histone, indicating their greater maximal stability [ $\Delta G(T_S)$ ]. Li et al. [40] have attempted to interpret the differences in thermodynamic stabilities of hyperthermophile rHMfB and mesophile rHFoB in terms of sequence and structural properties of the two histones. They have identified the amino acid residues responsible for the stability difference between the two histones using site-directed mutagenesis. Their results indicate that improved hydrophobic interactions in the histone dimer core, alleviation of electrostatic repulsion, and formation of additional ion pairs at the dimer surface are responsible for the higher thermostability of rHMfB.

Stability curves of cold-shock proteins from *T. maritima* (CspTm) [41], *B. subtilis* (CspBs) [42], and *E. coli* (CspA) [43] are plotted in figure 3. The stability curve of CspTm is both up- and right-shifted in comparison to those of CspA and CspBs. CspTm is a single-domain protein of 66 residues. It shows high sequence and structural identity with the mesophilic cold-shock proteins, CspA and CspBs. A three-dimensional model of CspTm was built by homology modeling using the CspBs X-ray structure as a template. The model structure indicated an increased number of surface salt bridges in CspTm [44]. In a series of related studies, Mueller et al. [45] have solved the X-ray structure of a cold-shock protein from the thermophile *B. caldolyticus* (CspBc) and compared it with that of CspBs. The distribution of surface charges was found to be different and overall favorable in CspBc. The amino acid sequences of CspBc and CspBs differ at 12 positions. Perl et al. [46] have produced 12 variants of CspBc, each containing an amino acid substitution in CspBc of the residue in the corresponding position in CspBs. They have reported that only two surface-exposed residues, Arg 3 and Leu 66, are responsible for the increase in stability of the thermophilic protein. Pace [47] has noted that a single mutation Arg 3 → Glu is capable

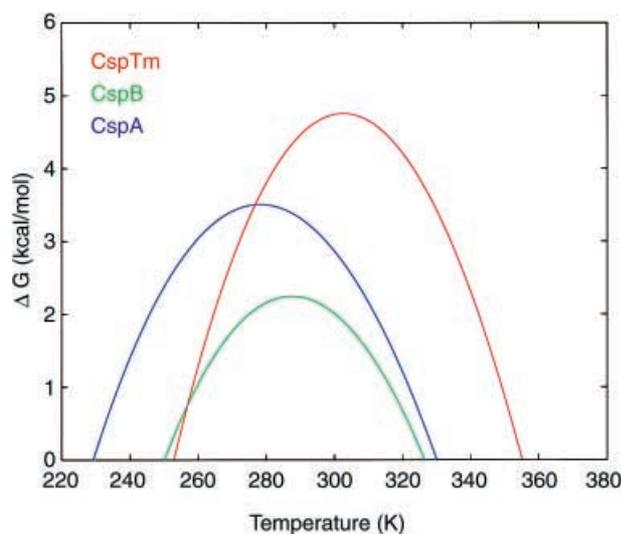


Figure 3. Comparison of protein stability curves for *E. coli*, *B. subtilis*, and *T. maritima* cold-shock protein in blue, green and red, respectively. The stability curve for the thermophilic *T. maritima* cold-shock protein is both up- and right-shifted in comparison to the stability curves for the mesophilic *E. coli* and *B. subtilis* cold-shock proteins [data from 41–43].

of turning the thermophilic CspBc into a mesophilic protein.

### Sequence and structural differences in thermophilic and mesophilic proteins

In the Introduction, we have listed different strategies adopted by thermophilic proteins to achieve stability and activity at high temperatures. A strategy that would lower the conformational entropy at high temperatures, strengthen the hydrophobic core, and optimize electrostatic interactions may impart thermostability to a protein. These principles are behind the various sequence and structural factors seen in thermostable proteins, with different thermophilic proteins adopting different single or combination of strategies. Interactions that impart thermostability are often characteristic of a protein family. Hence, comparison of a number of families containing homologous proteins from hyperthermophilic, thermophilic, and mesophilic organisms is a powerful method to understand the microscopic sequence/structural basis of protein thermostability. Furthermore, such comparisons may be useful for the rational design of thermostable proteins and in interpreting directed-evolution experiments [20, 48]. In the previous section we showed that the macroscopic thermodynamic differences among thermophiles and mesophiles are directly related to their microscopic properties. In this section, we review a few of these studies and our own work in this area.

Gromiha et al. [49] compared 48 physicochemical properties of individual amino acids in 16 families containing

homologous thermophilic and mesophilic proteins. They found that an increase in thermostability is correlated with the location of branch points in amino acids. For example,  $\beta$  and  $\gamma$ -branched amino acids increase protein thermostability. The Gibbs free energy of hydration,  $G_{\text{HN}}$ , for individual amino acids decreases in thermophilic proteins.

Reviews of the dominant structural factors in protein thermostability by Ladenstein and Antranikian [50] and Jaenicke and Bohm [51] have implicated electrostatic interactions such as salt bridge and hydrogen bond networks in protein thermostability. Optimization of atomic packing and hydrophobic interactions are also likely to be important [51]. However, Karshikoff and Ladenstein [24] have indicated that thermophilic and mesophilic proteins have similar atom packing. Szilagyi and Zavodszky [22] have compared 13 different structural features in 64 mesophilic and 29 thermophilic proteins representing 25 protein families. They also observed increased ion pair formation in thermophiles and hyperthermophiles. They further report that hyperthermophilic proteins are stabilized in a different way than the moderately thermophilic proteins. They observed differences in these two groups with respect to the number of ion pairs, atom packing, exposed polar surface area, and secondary-structure content [22].

Recently, we carried out a comprehensive analysis of sequence and structural factors enhancing protein thermostability [21]. We conducted a statistical examination of sequence and structural parameters in 18 non-redundant families of homologous thermophilic and mesophilic proteins. Each family consists of a thermophilic or a hyperthermophilic protein and its most similar (in both sequence and structure) mesophilic homolog. At the same time, each thermophilic protein is dissimilar in sequence and structure to other thermophilic proteins in our data set. This is also the case for the mesophilic proteins. Our goal was to construct a bias-free data set. The 18 non-redundant families used in our study are shown in table 1. High-resolution ( $R \leq 2.5 \text{ \AA}$ ) crystal structures for all proteins in our study are available in the Protein Data Bank (PDB) [52] and the relevant data are described in the literature [6, 12, 19, 23, 53–72]. We have compared a number of sequence and structural parameters among these families. These parameters include amino acid composition, proline substitution in loops,  $\alpha$ -helical content,  $\alpha$  helix geometry, hydrophobicity, compactness, polar/non-polar surface areas buried and exposed to water, insertion/deletions, oligomerization, hydrogen bonds and salt bridges. We have also compared our observations with an independent set of 165 non-redundant monomeric proteins.

Thermophilic protein sequences prefer Arg and Tyr and avoid Cys and Thr residues. Insertion/deletions and proline substitutions in loops do not show consistent trends

Table 1. Families of thermophilic and mesophilic proteins.

Protein family name	Thermophilic							Mesophilic							r. m. s. d. (Å)	Id (%)
	Organism	Stability	T <sub>L</sub> (°C)	PDB entry	Resolution (Å)	Oligomeric state	Nres	Organism	T <sub>L</sub> (°C)	PDB entry	Resolution (Å)	Oligomeric state	Nres			
Citrate synthase <sup>a</sup>	<i>Pyrococcus furiosus</i>	half-life of 170.0 min at 100 °C	100	1AJ8	1.9	dimer	741	chicken heart	37	1CSH	1.6	dimer	870	1.68	26.2	
Malate dehydrogenase <sup>b</sup>	<i>Thermus flavus</i>	fully active after 60 min at 90 °C	70–75	1BDM	2.5	dimer	644	pig	37	4MDH	2.5	dimer	666	0.94	54.1	
Rubredoxin <sup>c</sup>	<i>Pyrococcus furiosus</i>	stable for >24 h at 95 °C T <sub>m</sub> = 176 – 195 °C	100	1CAA	1.8	monomer	53	<i>Desulfovibrio vulgaris</i>	34–37	8RXN	1.0	monomer	52	0.69	66.7	
Cyclodextrin glucotransferase (CGTase) <sup>d</sup>	<i>Thermoanaerobacterium thermosulfurigenes</i>	>90% catalytic activity when kept at 80 °C for 5 h	60	1CIU	2.3	monomer	683	<i>Bacillus circulans</i>	30–40	1CDG	2.0	monomer	686	0.70	70.5	
EF-TU and EF-TU/TS complex <sup>e</sup>	<i>Thermus aquaticus</i>	temperature optimum ~70 °C	70–72	1EFT	2.5	monomer	405	<i>Escherichia coli</i>	37	1EFUC	2.5	A <sub>2</sub> B <sub>2</sub> tetramer (363 in chain C)	1290	1.5	57.6	
Glutamate dehydrogenase <sup>f</sup>	<i>Pyrococcus furiosus</i>	half-life of 12 h at 100 °C T <sub>m</sub> = 113 °C	75–105	IGTM	2.2	hexamer	2502	<i>Clostridium symbiosum</i>	30–37	1HRD	1.96	hexamer	2694	1.38	34.3	
Lactate dehydrogenase <sup>g</sup>	<i>Bacillus stearothermophilus</i>	active for 30 min at 80 °C	40–65	1LDN	2.5	tetramer	1264	<i>Plasmodium falciparum</i>	37	1LDG	1.74	tetramer	1260	1.25	28.4	
Thermolysin and neutral protease <sup>h</sup>	<i>Bacillus thermoproteolyticus</i>	50% activity after 1 h at 80 °C	52.5	1LNF	1.7	dimer	634	<i>Bacillus cereus</i>	30	1NPC	2.0	monomer	317	0.86	73.3	
3-Phosphoglycerate kinase (PGK) <sup>i</sup>	<i>Bacillus stearothermophilus</i>	T <sub>m</sub> = 67 °C	40–65	1PHP	1.65	monomer	394	<i>Saccharomyces cerevisiae</i>	25–30	1QPG	2.4	dimer	830	1.28	51.4	
Dimerization domain of EF-TS and EF-TU/TS complex <sup>j</sup>	<i>Thermus thermophilus</i>	does not denature up to 95 °C	70–75	1TFE	1.7	dimer	284	<i>Escherichia coli</i>	37	1EFU B	2.5	A <sub>2</sub> B <sub>2</sub> tetramer (282 in chain B)	1290	1.24	40.8	
CheY <sup>k</sup>	<i>Thermotoga maritima</i>	T <sub>m</sub> = 95 °C ΔH° = 78 kcal/mol optimum temperature = 90 °C	80–85	1TMY	1.9	monomer	118	<i>Escherichia coli</i>	37	3CHY	1.66	monomer	128	1.39	28.6	
Methionine aminopeptidase <sup>l</sup>	<i>Pyrococcus furiosus</i>	half-life of 4.5 h at 90 °C	100	1XGS	1.75	dimer	590	<i>Escherichia coli</i>	37	1MAT	2.4	monomer	263	1.39	30.6	

Table (continued)

Endo-1,4-b xylanase <sup>m</sup>	<i>Thermomyces lanuginosus</i>	highest activity at 65 °C for 15-min reaction	50	1YNA	1.55	mono- mer	193	<i>Bacillus circulans</i>	30–40	1XNB	1.49	mono- mer	185	1.14	50.9
Adenylate kinase <sup>a</sup>	<i>Bacillus stearother- mophilus</i>	T <sub>m</sub> = 74.5 °C ΔH = 145 kcal/mol	40–65	1ZIN	1.65	mono- mer	217	<i>Saccharomyces cerevisiae</i>	25–30	1AKY	1.63	mono- mer	218	1.22	42.0
Ferredoxin <sup>o</sup>	<i>Bacillus thermoproteo- lyticus</i>		52.5	2FXB	2.3	monomer	81	<i>Clostridium acidurici</i>	19-37	1FCA	1.8	mono- mer	55	1.27	24
Inorganic pyrophosphatase (hydrolase) <sup>p</sup>	<i>Thermus thermophilus</i>	retains 50% of initial activity after 1 h at 90 °C	70–75	2PRD	2.0	hexamer	1044	<i>Escherichia coli</i>	37	1INO	2.2	hexamer	1050	1.10	48.5
Manganese superoxide dismutase <sup>q</sup>	<i>Thermus thermophilus</i>		70–75	3MDS	1.8	tetramer	812	<i>Homo sapiens</i>	37	1QNM	2.3	tetramer	792	1.17	53.2
Phospho- fructokinase <sup>r</sup>	<i>Bacillus stearothermo- philus</i>		40–65	3PFK	2.4	tetramer	1276	<i>Escherichia coli</i>	37	2PFK	2.4	tetramer	1208	0.87	57.1

T<sub>L</sub>, living temperature; T<sub>m</sub>, melting temperature; Nres, number of residues in the whole protein; r.m.s.d., root mean square deviation; Id sequence identity. r.m.s.d. and Id were computed for individual chains in thermophilic and mesophilic proteins. Values here represent best matches. [Reprinted with permission from ref. 21; Oxford University Press, Oxford, ©2000.]

<sup>a</sup> Best match was obtained between chain B of 1AJ8 and 1CSH [12].

<sup>b</sup> Best match was obtained between chain B of 1BDM and chain B of 4MDH [53].

<sup>c</sup> There is more than one T<sub>m</sub> estimate for rubredoxin. That used here is from Hiller et al. [55].

<sup>d</sup> Knegt et al. [56].

<sup>e</sup> 1EFU corresponds to 1EFT and 1TFE in the thermophilic proteins. Best match for 1EFT was obtained with chain C of 1EFU [57].

<sup>f</sup> Best match was obtained between chain B of 1GTM and chain B of 1HRD [58]. Value of T<sub>m</sub> for 1GTM was obtained from Adams [59] and Klump et al. [60].

<sup>g</sup> Crystal asymmetric unit of 1LDN contains two copies of the molecule [61]. The first copy was used. Best match was obtained between chain C of 1LDN and 1LDG.

<sup>h</sup> Best match was obtained between chain E of 1LNF and 1NPC [6, 62]; activity data are from Singleton and Sainsbury [63].

<sup>i</sup> T<sub>m</sub> for mesophilic enzyme = 53 °C. ΔΔG ~ 5 kcal/mol [64, 65].

<sup>j</sup> Best match for 1TFE was obtained with chain B of 1EFU [66].

<sup>k</sup> Usher et al. [19].

<sup>l</sup> Best match was obtained between chain B of 1XGS and 1MAT [67].

<sup>m</sup> Data on activity were taken from Gomes et al. [68].

<sup>n</sup> T<sub>m</sub> for mesophilic adenylate kinase is 48 °C. ΔH<sub>m</sub> = 340 kJ/mol [69].

<sup>o</sup> Fukuyama et al. [70].

<sup>p</sup> Best match was obtained between the chains given in the asymmetric units of 2PRD and 1INO [23, 71].

<sup>q</sup> Asymmetric unit of 1QNM contains two identical chains of 198 residues each. Match was found to be best when both chains of 1QNM are simultaneously aligned with the chain in the asymmetric unit of 3MDS.

<sup>r</sup> Rypniewski and Evans [72].

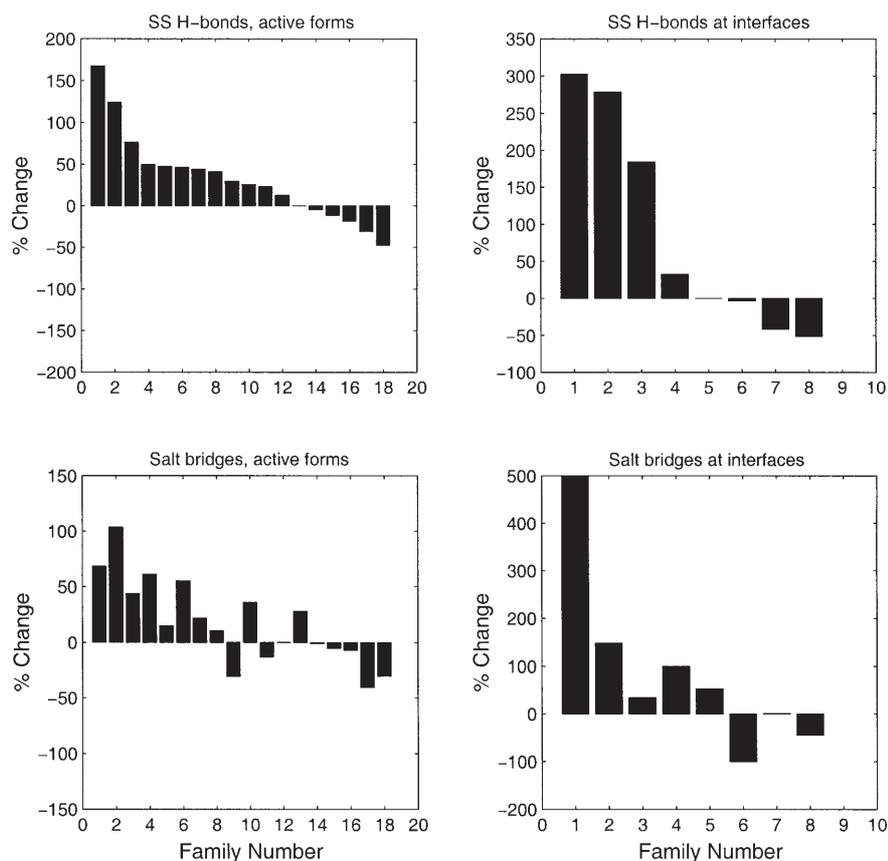


Figure 4. Plots depicting changes in side chain-side chain hydrogen bonds (SS H-bonds) and salt bridges in biochemically relevant forms of proteins and at interfaces in various families of thermophilic and mesophilic proteins. A positive change indicates that the thermophilic protein has a higher content compared to its mesophilic homologue, while a negative change indicates that the thermophilic protein has a lower content than its mesophilic homologue. For the majority of the families, SS H-bond and salt bridge content increases for thermophilic proteins. For each subplot, the X-axis denotes the family number while the Y-axis represents the percent change in the property indicated at the top of the subplot. The data on interfaces are available for eight families only. [Reprinted with permission from ref 21. Oxford University Press, Oxford, © 2000].

between thermophiles and mesophiles. Structural properties such as hydrophobicity, compactness, polar and non-polar contributions to surface areas, the number of main chain-main chain and main chain-side chain hydrogen bonds have similar values in thermophiles and mesophiles. Thermophilic proteins contain a greater fraction of residues in  $\alpha$ -helical conformation and thermophilic  $\alpha$  helices avoid Pro to a greater extent. However,  $\alpha$  helices in thermophilic and mesophilic proteins have similar overall geometries. Salt bridges and side chain-side chain hydrogen bonds increase in the majority of the thermophilic proteins, both within the subunits and at the interfaces (fig. 4). We have also investigated whether the changes in various structural properties like oligomerization, chain length (number of residues), hydrophobicity, compactness, hydrogen bonds and salt bridges are correlated with either the living temperatures of the thermophilic organisms or the melting temperatures of the thermophilic proteins. We observed a weak correlation between the increase in the number of salt bridges and the melting temperature of the thermophilic proteins.

### Genome-based studies on protein thermostability

Traditionally, studies such as those described above have been hampered by insufficient sequence and structural data. This situation is rapidly changing, owing to the availability of complete genome sequences of hyperthermophilic organisms and structural genomics efforts. In this section, we describe results obtained from analysis of complete genomic sequences of hyperthermophilic and thermophilic organisms.

As of January 2001, The Institute for Genomic research (TIGR) comprehensive microbial resource (cmr) web site ([www.tigr.org/tigr-scripts/CMR2/CMRGenomes.sp1](http://www.tigr.org/tigr-scripts/CMR2/CMRGenomes.sp1)) contained information on 37 complete microbial genomes. Nine of these genomes belong to thermophilic and hyperthermophilic organisms [73–81]. These organisms are listed in table 2 along with their optimum growth conditions. In addition, genome-sequencing projects of several thermophilic organisms including *Sulfolobus solfataricus*, *Pyrococcus furiosus*, *Pyrobaculum aerophilum*, *Thermospora fusca*, *Thermus flavus* and *T. thermophi-*

Table 2. Thermophilic organisms for which complete genome sequences are available.

Organism name	Kingdom	Optimum growth conditions	Reference(s)
<i>Aeropyrum pernix</i> K1	archaea	90–95 °C pH 7.0 salinity 3.5%	Sako et al. [73]
<i>Archaeoglobus fulgidus</i> DSM4304	archaea	83 °C	Klenk et al. [74]
<i>Methanobacterium thermoautotrophicum</i> delta H	archaea	65 °C	Smith et al. [75]
<i>Methanococcus jannaschii</i> DSM 2661	archaea	85 °C >200 atm pressure	Bult et al. [76]
<i>Pyrococcus horikoshii</i> (shinkaj) OT3	archaea	98 °C	Kawarabayasi et al. [77]
<i>Pyrococcus abyssi</i> GE5	archaea	103 °C 200 atm pressure	Not available
<i>Thermoplasma acidophilum</i> DSM 1728	archaea	55–60 °C pH 0.5–4	Ruepp et al. [78]
<i>Aquifex aeolicus</i> VF5	bacteria	96 °C	Deckert et al. [79]
<i>Thermotoga maritima</i> MSB8	bacteria	80 °C	Nelson et al. [80]
<i>Thermoplasma volcanium</i> <sup>a</sup>	archaea	60 °C	Kawashima et al. [81]

Data complete to mid-January 2001. Information taken from TGR comprehensive microbial resource internet page.

<sup>a</sup> Report found during pubmed search.

lus are in various stages of completion [82]. Literature reports on completed genomes do not indicate any specific gene loci for thermophilicity although some hyperthermophilic organisms do contain heat-shock proteins which act as chaperones.

Chakravarty and Vardarajan [83] have compared amino acid sequences of soluble proteins in complete genomes of 8 thermophilic and 12 mesophilic organisms. On average, thermophilic proteins were found to contain fewer residues ( $268 \pm 38$ ) than the mesophilic proteins ( $310 \pm 16$ ). The increase in proportion of charged residues (Arg, Lys, His, Asp, Glu) and decrease in proportion of uncharged polar residues (Ser, Thr, Gln, Asn, Cys) in thermophilic proteins are statistically significant. Hydrophobic  $\beta$ -branched residues were also found to be more frequent in the thermophilic proteins. Thompson and Eisenberg [84] have also observed a statistically significant trend for shorter sequence lengths in thermophilic proteins in a comparison involving the complete genomes of 20 mesophilic, thermophilic, and hyperthermophilic or-

ganisms. They have observed that the sequence deletion sites in thermophilic proteins correspond to the exposed loop regions. Hence, they concluded that deletion of exposed loops is a natural mechanism for enhancing protein thermostability, in addition to improving electrostatic interactions. Further, they have observed that thermophilic proteins contain more glutamate, valine, arginine, and glycine residues and less glutamine, serine, asparagine, and lysine residues.

Zhang [85] has computed the mean lengths for proteins in complete genomes of 22 organisms (5 archaea, 15 bacteria, and 2 eukaryotes). He found that the mean protein lengths in the archaeal, bacterial, and eukaryotic species have averages of  $270 \pm 9$ ,  $330 \pm 5$ , and  $449 \pm 25$ , respectively. Hence, the differences in protein lengths may also have an additional biological purpose.

Cambillau and Claverie [86] have compared the composition of proteins from 30 complete genomes (22 mesophiles, 1 thermophile, and 7 hyperthermophiles). They have observed a statistically significant increase in the proportion of charged residues (Lys, Arg, Asp, Glu) for proteins in the hyperthermophiles and in the thermophile. A decrease in the proportion of uncharged polar residues (Asn, Gln, Ser, Thr) was also statistically significant. They also computed water-accessible surface areas for each amino acid in 131 mesophilic proteins and 58 hyperthermophilic proteins. The hyperthermophilic proteins have a greater proportion of charged residues at the surface. The authors argued that higher thermal stability of the hyperthermophilic proteins is imprinted in the genomes of the hyperthermophilic organisms. An increase in charged residues at the expense of uncharged polar residues was also reported by Haney et al. [87] based on a comparison of amino acid sequences of 115 proteins from the hyperthermophilic archaeon *Methanococcus jannaschii* with their homologs from mesophilic *Methanococcus* species.

### Protein thermostability and electrostatic interactions

While there may be several strategies to attain thermostability, nature appears to have used an improvement in electrostatic interactions most frequently. Hence, we devote a separate section to the role of electrostatics in protein thermostability. Electrostatic interactions such as salt bridges and their networks have important roles in protein folding, structure, and function [88–91]. Intuitively, electrostatic interactions should stabilize folded proteins. If this were always true, one could easily rationalize the increase in electrostatic interactions in the thermophilic proteins. Instead, estimates of the energetic contribution of salt bridges vary from being stabilizing through being insignificant or small to even being destabilizing [88, 89, 92–97]. Recently, Maves and Sligar [98] performed com-

binatorial mutagenesis experiments to study the high thermal stability of CYP-119, a cytochrome P450, from *S. solfataricus*. They generated a random library of point mutants and screened for variants that are less thermostable than the wild-type CYP-119. In total, they identified 13 mutants, 11 of them point mutants. Most of the mutations in the CYP-119 were changes of surface charged residues that are involved in salt bridge formation. Merz et al. [99] mutated two salt bridges in *T. maritima* indoleglycerol phosphate synthase (tIGPS). One of the salt bridges fixes the N terminus to the protein core and the second is an interhelical salt bridge. Both salt bridges stabilize tIGPS. Strop and Mayo [100] estimated the contribution of two surface salt bridges in *P. furiosus* rubredoxin variant (PFRD-XC4) using double-mutant cycles. They found one of the salt bridges to be stabilizing and the other to be slightly destabilizing. Using site-directed mutagenesis, Kawamura et al. [101] have observed that disruption of the salt bridge Glu 34-Lys 38 in the DNA-binding protein HU from *B. stearothermophilus* reduces its thermostability. A correlation between salt bridge networks and melting temperature has been observed for the D-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) family [102]. Yip et al. [103] have also observed a similar correlation between salt bridge networks and thermostabilities of glutamate dehydrogenases from several hyperthermophilic, thermophilic, and mesophilic sources. However, Lebbink et al. [104] could not improve the stability of *T. maritima* glutamate dehydrogenase by introducing a six-residue ion pair network in the hinge region, despite the anticipated improvement based on a comparison between homologous three-dimensional structures of *P. furiosus* and *T. maritima* glutamate dehydrogenases. Grimsley et al. [105] have shown that the stability of Rnase T1 can be increased by improving long-range electrostatic interactions among charged groups on the protein surface.

At the same time, several theoretical calculations have suggested that the net electrostatic contribution to the free energy of protein folding and binding is destabilizing [106, 107]. This was explained as follows. In the unfolded state, the charged residues of a protein are fully solvated by water molecules. Solvation of charged residues is energetically favorable. As the protein folds, these solvated charged residues must desolvate. Hence, depending upon their location in the protein, each charged residue incurs an energetically unfavorable desolvation penalty. The penalties incurred by the charged residues in a folded protein often remain uncompensated. Based on analysis of 21 salt bridges from nine proteins, Hendsch and Tidor [106] concluded that most salt bridges in proteins are destabilizing. However, de Bakker et al. [108] performed molecular dynamic simulations on Sac7d from the hyperthermophile *S. acidocaldarius* at 300, 360, and 550 K. They concluded that the salt bridges

contribute favorably toward protein stability at elevated temperatures.

The above observations lead to an apparent ambiguity. On one hand, an increase in the proportion of charged residues and improved electrostatic interactions appear to be most consistent among the factors enhancing protein thermostability. On the other hand, electrostatic interactions such as salt bridges and their networks may destabilize the folded/bound states of the proteins in some (if not all) cases. Hydration free energies of amino acids change with temperature. This is due to a decrease in the dielectric constant of water and the contribution of entropic effects at higher temperatures [109, 110]. Studies by Elcock and coworkers indicate a reduced energy penalty for desolvation of charged residues in a folded protein at elevated temperatures. Hence, salt bridges in thermophiles and hyperthermophiles may be stabilizing. Xiao and Honig [111] have computed the electrostatic contributions to protein stability for four hyperthermophilic proteins and their mesophilic homologs. They observed that the hyperthermophilic proteins have greater electrostatic contributions than their mesophilic homologs. They conclude that optimization of electrostatic interactions in the hyperthermophilic proteins results in their greater electrostatic contribution.

Recently, we carried out extensive analyses of electrostatic interactions in proteins [88–91]. Theoretical modeling of electrostatic properties of a protein in aqueous solution requires an accurate description of the solute (protein), the solvent (water), and their interaction. Continuum electrostatic methods are based on classical electrostatics. In these methods, the solvent is treated as a homogeneous continuum and only the bulk (average) properties, such as dielectric constant, are taken into account. Hence, there are no explicit water molecules present around the protein in these calculations. However, full atomic details (as in X-ray crystal structure) of the protein molecules are utilized. These methods are widely used as quantitative tools. A method for computing the electrostatic contributions to the free energy change upon salt bridge formation, using continuum electrostatic calculations, was reported by Hendsch and Tidor [106]. We refer to the electrostatic contribution to the free energy change upon salt bridge formation in a protein as the electrostatic strength of the salt bridge. The method of Hendsch and Tidor calculates the electrostatic strength of a salt bridge relative to computer-generated mutations of the salt-bridging residue sidechains to their hydrophobic isosteres. The hydrophobic isosteres are the salt-bridging residue side chains with their partial atomic charges set to zero. The electrostatic strength of a salt bridge can be partitioned into three component terms according to the following equation:

$$\Delta\Delta G_{\text{tot}} = \Delta\Delta G_{\text{dslv}} + \Delta\Delta G_{\text{brd}} + \Delta\Delta G_{\text{prt}} \quad (11)$$

where  $\Delta\Delta G_{\text{dsiv}}$  is the sum of the unfavorable desolvation penalties incurred by the individual salt-bridging residues in the folded protein.  $\Delta\Delta G_{\text{brd}}$  is the favorable bridge energy due to the electrostatic interaction of the side chain charged groups with each other in the protein.  $\Delta\Delta G_{\text{prt}}$  represents the electrostatic interaction of the salt-bridging side chains with the charges in the rest of the protein.

Using this method, we have computed the electrostatic strengths of 222 non-equivalent salt bridges derived from 36 non-homologous high-resolution monomeric protein crystal structures [89]. We found most of the salt bridges in our data set to be stabilizing. The electrostatic strength of a salt bridge depends upon three factors, namely the geometry of the salt bridge, the location of the salt-bridging residues and the interaction of the salt-bridging residues with other charged residues in the protein. The geometry of a salt bridge is a critical factor in determining its electrostatic strength. Salt bridges with favorable geometrical positioning of the interacting side chain charged groups are likely to stabilize the protein structure. These observations have gained further support from our recent work in this direction [90, 91, Kumar and Nussinov, unpublished data].

We have further used this method to analyze salt bridges in glutamate dehydrogenase from a hyperthermophile (*P. furiosus*) and a mesophile (*Clostridium symbiosum*) [88]. Glutamate dehydrogenase (GDH) catalyzes reversible oxidative deamination of L-glutamate to 2-oxoglutarate

and ammonia using  $\text{NAD}^+$  or  $\text{NADP}^+$  as cofactor. *P. furiosus* GDH (PfGDH) is extremely thermostable, its melting temperature being  $113^\circ\text{C}$  (table 1). The mesophilic *C. symbiosum* GDH (CsGDH) shares 34% sequence identity with PfGDH. In both organisms, the biochemically active GDH is a homohexamer. The three-dimensional structures for both GDHs are highly similar (table 1). In contrast to PfGDH, CsGDH has a half-life of only 20 min at  $52^\circ\text{C}$  and its melting temperature is  $55^\circ\text{C}$ . Our previous analysis [21] and a crystallographic analysis [58] have found an increase in salt bridge formation in PfGDH. Salt bridges in PfGDH show an  $\sim 70\%$  increase for the whole hexameric biological unit compared to CsGDH. The large ( $60^\circ\text{C}$ ) difference in melting temperatures of PfGDH and CsGDH but their high sequence and structural similarity make this thermophile-mesophile pair a good model for investigating the molecular basis of thermostability.

Figure 5 shows the salt-bridge-forming residues in the corresponding monomers of PfGDH (B chain of 1GTM; PDB [52] file for PfGDH) and CsGDH (the B chain of 1HRD; PDB [52] file for CsGDH). The PfGDH monomer has several salt bridges near the active site which are missing in CsGDH. We have computed the electrostatic strengths for 29 (out of 40) salt bridges in a PfGDH monomer. We have also computed the electrostatic strengths of 17 (out of 20) salt bridges in the corresponding CsGDH monomer [88].

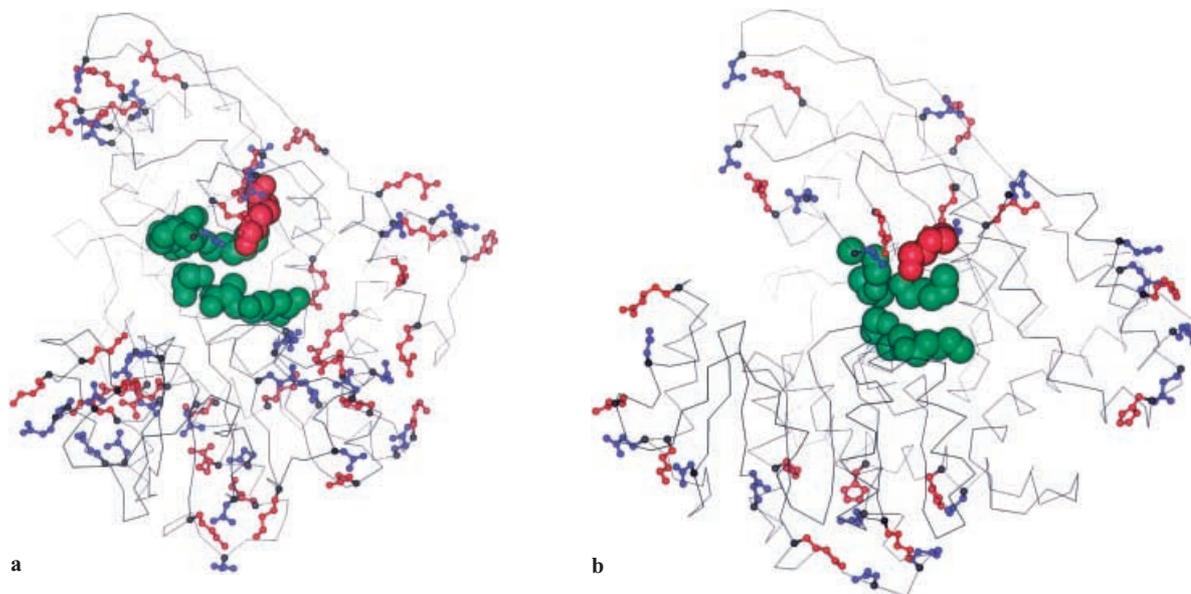


Figure 5. Diagrams showing  $\text{C}^\alpha$  trace, active-site and salt-bridge-forming residues in a subunit of glutamate dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus* (chain B in 1GTM) (a) and mesophilic Gram-positive bacterium *Clostridium symbiosum* (chain B in 1HRD) (b). All the heavy atoms are shown for salt-bridge-forming as well as for active-site residues. Active-site residues are shown with CPK representation. The conserved active site Lys which participates in salt bridge network formation as well as in enzyme activity is shown in red. Other active-site residues are in green. Residues with side chain atoms shown in blue are the positively charged residues (Arg, Lys, His), while the residues with side chain atoms shown in red are the negatively charged residues (Glu, Asp), in ball-and-stick representation.  $\text{C}^\alpha$  atoms of the salt-bridge-forming residues are shown in black. Thermophilic glutamate dehydrogenase has several additional salt bridges in the neighborhood of the active site compared to the mesophilic glutamate dehydrogenase. [Reprinted with permission from ref. 88. Wiley-Liss, New York, © 2000].

Table 3 presents the electrostatic strengths of the salt bridges in CsGDH and PfGDH monomers. The results are shown (in terms of kcal/mol) at the respective optimum growth temperatures for *P. furiosus* and *C. symbiosum*. The solvent (water) dielectric constant was 80.0 in these calculations. The PfGDH monomer has two destabilizing salt bridges while the CsGDH monomer has six. The unfavorable desolvation energy penalty ( $\Delta\Delta G_{\text{dsolv}}$ ) is compensated by the bridge ( $\Delta\Delta G_{\text{brd}}$ ) and protein ( $\Delta\Delta G_{\text{prt}}$ ) energy terms only marginally for salt bridges in the CsGDH monomer. In contrast, the desolvation penalty is significantly compensated by the bridge and protein energy terms in the salt bridges of the PfGDH monomer (average total free energy change =  $-5.5$  kcal/mol). At  $100^\circ\text{C}$ , the living temperature for *P. furiosus*, the dielectric constant of water falls to 55.51. Table 3 also shows the electrostatic strengths of the salt bridges in PfGDH with a solvent dielectric constant of 55.51. The salt bridges become more stabilizing, with the average  $\Delta\Delta G_{\text{tot}}$  decreasing from  $-5.52$  kcal/mol to  $-6.49$  kcal/mol.

Why are the salt bridges in the PfGDH monomer highly stabilizing? The answer may lie in the formation of salt

bridge networks and the cooperative nature of the electrostatic interactions. Table 3 shows that for salt bridges in the CsGDH monomer, the bridge energy term ( $\Delta\Delta G_{\text{brd}}$ ) is considerably larger in magnitude than the protein energy term ( $\Delta\Delta G_{\text{prt}}$ ). This indicates that the interaction of the salt-bridging residue side chains with each other is stronger than the interaction of these side chains with the charges in the rest of the CsGDH. That is, salt bridges in CsGDH tend to be isolated. In contrast,  $\Delta\Delta G_{\text{brd}}$  and  $\Delta\Delta G_{\text{prt}}$  have similar magnitudes in the PfGDH monomer (table 3). Hence, in this case, the interactions of charged side chains in the salt-bridge-forming residues with the rest of the protein are almost as significant as the interaction of these side chains with each other. PfGDH is particularly rich in salt bridge networks [21, 58]. There are eight clusters of salt bridge networks in the B chain of 1GTM and several of these clusters lie around the active site (fig. 5b). In total, these clusters account for 23 out of 29 salt bridges. These calculations illustrate the cooperative nature of electrostatic interactions, indicating that optimization of these interactions can lead to increased protein thermal stability.

Table 3. Energies of salt bridges in GDH of *Pyrococcus furiosus* (Pf) and *Clostridium symbiosum*.

Salt Bridge	ASA <sub>av</sub> (%)	$\Delta\Delta G_{\text{tot}}$ (kcal/mol)	$\Delta\Delta G_{\text{dsolv}}$ (kcal/mol)	$\Delta\Delta G_{\text{brd}}$ (kcal/mol)	$\Delta\Delta G_{\text{prt}}$ (kcal/mol)
Energies of salt bridges in CsGDH(B chain of 1HRD) <sup>a</sup>					
R6-E10	42.8	+0.099	2.986	-1.139	-1.749
R6-E43	40.4	-1.144	3.162	-2.059	-2.247
E18-K104	7.5	+1.514	10.751	-7.346	-1.892
H39-E 41	25.6	-1.387	6.075	-2.909	-4.554
R78-D160	0.0	-6.588	26.376	-28.380	-4.584
R93-D165	25.9	-7.290	6.848	-9.961	-4.177
K125-D165	18.3	-6.268	9.586	-5.165	-10.689
D137-R171	45.7	-0.311	4.935	-2.284	-2.962
R171-E172	22.4	-7.107	7.113	-6.249	-7.971
E218-H403	22.4	+1.504	4.368	-3.786	+0.472
E224-K340	25.3	+3.561	5.677	-2.599	+0.482
D226-K231	24.7	-6.559	5.085	-5.350	-6.294
K248-E251	34.4	+0.832	6.049	-2.359	-2.858
D268-K277	4.7	-1.618	15.803	-14.001	-3.421
E276-K298	52.7	-1.621	1.459	-1.783	-1.297
R289-D294	33.6	-0.356	7.765	-2.573	-5.548
H410-D411	24.5	+7.072	9.694	-4.092	+1.470
Average	26.5 ± 14.3	-1.536 ± 4.068	7.867 ± 5.856	-6.002 ± 6.657	-2.734 ± 2.598
Average at 25 °C	26.5 ± 14.3	-1.476 ± 3.910	7.560 ± 5.628	-5.768 ± 6.398	-2.627 ± 2.497
Energies of salt bridges in PfGDH (B chain of 1GTM) <sup>a</sup>					
E7-K11	50.6	+2.395	3.970	-1.292	-0.283
R15-D397	20.8	-6.131	7.300	-9.805	-3.627
R57-D139	0.0	-7.357	32.101	-34.096	-5.362
R72-E77	7.2	-8.931	16.168	-22.856	-2.243
R72-D144	22.9	-19.098	12.722	-4.678	-27.142
D97-K379	38.1	-2.459	5.646	-4.492	-3.613
K104-D144	19.3	-4.199	9.661	-6.636	-7.224
E188-R192	7.1	-15.086	9.965	-13.464	-11.587
E188-R370	19.2	-13.107	9.592	-14.524	-8.176
R192-D234	32.0	-1.384	8.920	-2.947	-7.358
R199-E200	10.8	-8.110	13.040	-8.922	-12.227

Table 3 (continued)

Salt Bridge	ASA <sub>av</sub> (%)	$\Delta\Delta G_{\text{tot}}$ (kcal/mol)	$\Delta\Delta G_{\text{dsiv}}$ (kcal/mol)	$\Delta\Delta G_{\text{brd}}$ (kcal/mol)	$\Delta\Delta G_{\text{prt}}$ (kcal/mol)
R199-D374	21.4	-4.690	9.588	-9.506	-4.722
E200-K203	27.0	-6.004	8.661	-8.443	-6.221
D208-K213	52.6	-5.265	5.160	-6.034	-4.391
K229-E233	41.4	-10.042	4.950	-9.834	-5.157
K229-D258	25.7	-5.927	7.887	-5.714	-8.100
D244-K264	13.0	-6.278	16.938	-6.974	-16.241
E259-K262	47.5	-1.972	0.108	-2.099	+0.019
K271-D272	48.1	-4.535	0.921	-5.418	-0.038
D290-K312	31.2	-1.939	6.265	-2.677	-5.526
D307-K333	50.2	-3.703	2.659	-2.311	-4.051
E316-R396	4.4	-6.729	20.492	-16.091	-11.131
D327-H394	6.5	+10.849	17.637	-6.711	-0.077
D327-R396	7.6	-10.537	15.574	-11.559	-14.553
D340-R396	5.3	-13.865	16.281	-4.030	-26.116
E371-K375	54.0	-1.093	2.233	-2.539	-0.787
K379-D383	32.3	-2.055	7.333	-4.747	-4.641
D383-R406	19.4	-1.507	9.382	-6.307	-4.582
E390-K391	55.6	-1.792	1.130	-2.527	-0.396
Average	26.6 ± 17.5	-5.524 ± 5.747	9.731 ± 6.928	-8.162 ± 6.977	-7.093 ± 6.917
Average at 25 °C	26.6 ± 17.5	-4.413 ± 4.591	7.774 ± 5.535	-6.521 ± 5.574	-5.667 ± 5.526
Energies of salt bridges in B chain of 1GTM with water dielectric constant, $\epsilon = 55.51$ at 100 °C					
E7-K11	50.6	+2.027	4.060	-1.725	-0.307
R15-D397	20.8	-7.197	7.669	-10.830	-4.067
R57-D139	0.0	-8.198	31.304	-34.452	-5.049
R72-E77	7.2	-9.861	16.082	-23.665	-2.278
R72-D144	22.9	-20.748	12.983	-5.434	-28.296
D97-K379	38.1	-3.403	6.191	-5.559	-4.035
K104-D144	19.3	-5.187	10.301	-7.858	-7.630
E188-R192	7.1	-16.656	10.243	-14.348	-12.551
E188-R370	19.2	-14.382	9.746	-15.425	-8.703
R192-D234	32.0	-1.831	9.170	-3.749	-7.252
R199-E200	10.8	-9.601	13.055	-9.702	-12.954
R199-D374	21.4	-5.628	9.792	-10.443	-4.978
E200-K203	27.0	-6.980	8.817	-9.524	-6.273
D208-K213	52.6	-6.096	5.341	-6.907	-4.530
K229-E233	41.4	-11.076	5.124	-10.877	-5.323
K229-D258	25.7	-7.911	7.965	-7.302	-8.574
D244-K264	13.0	-7.102	16.855	-7.716	-16.240
E259-K262	47.5	-2.512	0.229	-2.758	+0.017
K271-D272	48.1	-4.535	0.921	-5.418	-0.038
D290-K312	31.2	-2.433	6.600	-3.370	-5.663
D307-K333	50.2	-4.589	2.804	-2.840	-4.553
E316-R396	4.4	-7.751	20.787	-17.100	-11.438
D327-H394	6.5	+9.967	17.887	-7.534	-0.386
D327-R396	7.6	-11.781	15.899	-12.759	-14.922
D340-R396	5.3	-14.999	16.829	-4.986	-26.842
E371-K375	54.0	-1.765	2.495	-3.292	-0.968
K379-D383	32.3	-3.144	7.764	-5.751	-5.157
D383-R406	19.4	-2.362	9.786	-7.178	-4.971
E390-K391	55.6	-2.419	1.280	-3.205	-0.494
Average	26.6 ± 17.5	-6.488 ± 5.985	9.931 ± 6.818	-9.024 ± 6.962	-7.395 ± 7.121

ASA<sub>av</sub> indicates the average accessible surface area of the two residues forming a salt bridge in the hexameric state of GDH.  $\Delta\Delta G_{\text{tot}}$  refers to the total electrostatic free energy of the salt bridge.  $\Delta\Delta G_{\text{dsiv}}$  indicates the desolvation energy penalty incurred by the salt bridge.  $\Delta\Delta G_{\text{brd}}$  is the free energy of the interaction of salt-bridge-forming side chains with each other.  $\Delta\Delta G_{\text{prt}}$  is the free energy of the interaction of salt-bridge-forming side chains with the rest of the protein. DELPHI calculations for 1HRD were performed with a water dielectric constant of 80, while those for 1GTM were performed with water dielectric constants of 80 and 55.51. The dielectric constant of water is 55.51 at 100 °C. Averages of various energy terms at room temperature are given for the purpose of comparison.

<sup>a</sup> Salt bridged whose residues have identical or nearly identical AsAs in mimomeic and hexameric states.

[Reprinted with permission from ref. 88. Wiley-Liss, New York, © 2000].

## Discussion and concluding remarks

How then do thermophilic proteins deal with heat? The forces that act to keep hyperthermophilic and thermophilic proteins functional are apparently similar to those in mesophilic proteins. Apart from the chaperoning activity of heat shock proteins and production of a few compatible solutes (see Sterner and Liebl [112] for a comprehensive account of these aspects of protein thermostability), the proteins in a hyperthermophilic cell take advantage of variations in their sequences and structures to deal with the extra heat in their environment. A number of strategies have been adopted by the proteins to improve their thermostability, although in a non-consistent way. Thus, for example, the well-known stabilizing factors in protein conformations, such as hydrophobicity, atomic packing, and main chain-main chain hydrogen bonds, although adopted by some hyperthermophilic proteins, do not show consistent, substantial variations between mesophiles, thermophiles, and hyperthermophiles [21, 24]. Protein stability also relates to function, and to the living temperature of the organism. Homologous thermophilic and mesophilic proteins have similar stabilities [ $\Delta G(T_L)$  at their respective organism optimum growth temperatures [Kumar et al., unpublished results]. When a mesophilic protein is exposed to higher temperatures, its stability decreases in accordance with its stability curve. The sequence and structural variations keep the stability of the homologous thermophilic protein roughly unchanged, enabling it to function properly at the elevated temperatures.

At elevated temperatures, proteins may be expected to have increased conformational disorder (entropy) because of greater atomic mobility (vibrations). There is also greater disorder in protein-solvent interactions. Chemical reaction rates increase and the stabilities of the substrate molecules decrease at such temperatures. In this scenario, even small conformational changes at or nearby active sites may affect catalysis. Hyperthermophilic and thermophilic proteins need a control mechanism to resist deleterious changes in their structures. Its absence may lead to loss of protein function due to structural deformation (and/or denaturation). Such a mechanism is likely to act in a delocalized manner throughout the protein structure to maintain its integrity at high temperatures.

The control mechanism may work by first strengthening the hydrophobic core, as observed in the increased occurrence of hydrophobic amino acids with branched side chains in the thermophilic proteins [49, 83, 84]. The extra hydrophobic interactions may help increase the enthalpy change at the melting temperature ( $\Delta H_G$ ). Second, surface loops are reduced or eliminated. Surface loops are usually mobile. Deletion of the loops reduces the conformational entropy of the hyperthermophilic proteins [84]. Furthermore, hyperthermophilic proteins tend to be

shorter [83, 84]. Smaller proteins have a smaller heat capacity change ( $\Delta C_p$ ) [33], leading to a broader protein stability curve and higher melting temperature. Third, the structural plasticity of hyperthermophilic and thermophilic proteins can be reduced by increasing and optimizing electrostatic interactions. A statistically significant increase in the proportion of charged residues at the expense of uncharged polar residues has been observed in hyperthermophilic proteins compared to mesophilic ones [e.g., 21, 22, 83, 86]. An increase in the proportion of charged residues leads to a stronger electrostatic effect in these proteins [21, 22, 88, 109–111]. An increase in stabilizing electrostatic interactions in and around the active site, ligand-binding site, or metal-binding site may help maintain the integrity of these sites at elevated temperatures [88]. In several proteins, these sites are buried in deep hydrophobic pockets. Burying charged residues in protein cores may not appear to be an attractive proposition due to the large desolvation penalties. However, buried charges occur in globular proteins at greater frequency than previously thought [113]. In our investigation of 222 salt bridges from 36 monomeric proteins [89], approximately one-third of the salt bridges were buried. Buried salt bridges were found to be more stabilizing than the surface-exposed ones. This is because electrostatic interactions both between the salt-bridging side chains, and between salt bridges and the charges in the rest of the protein are also stronger in the protein core due to the absence of solvent screening. The buried salt bridges and their networks around the active site in the PfGDH monomer are highly stabilizing [88]. Electrostatic interactions on the protein surface are also stronger in thermostable proteins, the outcome of a decrease in the dielectric constant of water to 55.51 at 100°C. Interaction of surface charged residues with solvent water molecules may oppose disorder in protein-solvent interactions at high temperature. In the folded state of a hyperthermophilic protein, the presence of salt bridges and their networks may provide kinetic barriers against protein unfolding [114]. These additional electrostatic interactions further increase the enthalpy change between the folded and unfolded states at the melting temperature.

Increased and improved electrostatic interactions should not be taken to mean that hyperthermophilic proteins are more rigid than their mesophilic homologs. These proteins are flexible at their optimum temperatures [115]. In the native state, a protein consists of an ensemble of conformers. The populations in the protein ensemble shift with changes in the protein environment such as temperature, pH, and presence/absence of ligands/cofactors [116–118]. Protein flexibility is essential for function. Our recent investigations into the role of electrostatic interactions in systemic protein flexibility [90, 91] have shown that salt bridges and their networks observed in protein crystal structures may easily break and reform in

solution. Furthermore, alternative electrostatic interactions, not seen in the crystal structures, may also often form and break in solution. Hence, an increase or decrease in electrostatic interactions may not directly imply an increase or decrease in protein rigidity.

Here, we have emphasized the role of electrostatic interactions in protein thermostability. However, proteins may also adopt other routes to thermostability. For example, the dodecameric state of ornithine carbamoyltransferase from *P. furiosus* is stabilized by hydrophobic interactions at the trimeric catalytic motif interfaces [119]. GDH from the same organism is stabilized by electrostatic interactions [58, 88]. Hence, thermostabilization strategies act at the molecular rather than the organismal level. Sequence and structural differences elucidated in families of homologous thermophilic and mesophilic proteins were also noted by Vanhove et al. [120] while comparing sequence and structural properties of five class A  $\beta$ -lactamases produced by various mesophilic bacteria.

### Future directions

The availability of larger sequence and structural data on proteins would undoubtedly lead to better understanding of protein stability. While our understanding of factors enhancing protein thermostability has improved considerably in recent years, rational design of thermostable proteins still remains an elusive goal. Currently, alleviation of electrostatic repulsion [105, 121, 122] and incorporation of electrostatic interactions with favorable geometries appear favorable design strategies, along with strong hydrophobic cores, better packing, shorter loops, choice of preferred residues, as well as several additional sequence and structural factors.

*Acknowledgement.* We thank Drs. Buyong Ba, Chung-jung Tsai, Neeti Sinha, Yuk Yin Sham and, in particular, Jacob V. Maizel for numerous helpful discussions. The research of R. Nussinov in Israel was supported in part by grant number 95-00208 from BSF, Israel, by a grant from the Israel Science Foundation administered by the Israel Academy of Sciences, by a Magnet grant, by a Ministry of Science grant, by the Tel Aviv University Basic Research grants and by the Center of Excellence, administered by the Israel Academy of Sciences. This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number NO1-CO-56000. The content of this publication does not necessarily reflect the view or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the U.S. Government.

- 1 Huber R., Huber H. and Stretter K. O. (2000) Towards the ecology of hyperthermophiles: biotops, new isolation strategies and novel metabolic properties. *FEMS Microbiol. Rev.* **24**: 615–623
- 2 Vihinen M. and Mantsala P. (1989) Microbial amylolytic enzymes. *Crit. Rev. Biochem. Mol. Biol.* **24**: 329–418
- 3 Adams M. W. W. and Kelly R. M. (1995) Enzymes from microorganisms in extreme environments. *Chem. Eng. News* **73**: 32–42
- 4 Persidis A. (1998) Extremophiles: unusual and useful molecules are found in life on the edge of environmental tolerance. *N. Biotechnol.* **16**: 593–594
- 5 Lehmann M., Pasamontes L., Lassen S. F. and Wyss M. (2000) The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta* **1543**: 408–415
- 6 Matthews B. W., Weaver L. H. and Kester W. R. (1974) The conformation of thermolysin. *J. Biol. Chem.* **249**: 8030–8044
- 7 Perutz M. and Raidt H. (1975) Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature* **255**: 256–259
- 8 Argos P., Rossmann M. G., Grau U. M., Zuber H., Frank G. and Tratschin J. D. (1979) Thermal stability and protein structure. *Biochemistry* **25**: 5698–5703
- 9 Zuber H. (1988) Temperature adaptation of lactate dehydrogenase: structural, functional and genetic aspects. *Biophys. Chem.* **29**: 171–179
- 10 Nosoh Y. and Skeiguchi T. (1990) Protein engineering for thermostability. *Trends Biotechnol.* **8**: 16–20
- 11 Gupta M. (1993) *Thermostability of Enzymes*, Springer, Berlin
- 12 Russell R. J. M., Ferguson J. M. C., Haugh D. W., Danson M. J. and Taylor G. L. (1997) The crystal structure of citrate synthase from the hyperthermophilic bacterium *Pyrococcus furiosus* at 1.9 Å resolution. *Biochemistry* **36**: 9983–9994
- 13 Warren G. L. and Petsko G. A. (1995) Composition analysis of  $\alpha$ -helices in thermophilic organisms. *Protein Eng.* **8**: 905–913
- 14 Rees D. C. and Adams M. W. W. (1995) Hyperthermophiles: taking heat and loving it. *Structure* **3**: 251–254
- 15 Querol E., Perez-Pons J. A. and Mozo-Villarias A. (1996) Analysis of protein conformational characteristics related to thermostability. *Protein Eng.* **9**: 256–271
- 16 Jaenicke R. (1996) Glyceraldehyde-3-phosphate dehydrogenase from *Thermotoga maritima*: strategies of protein stabilization. *FEMS Microbiol. Rev.* **18**: 215–224
- 17 Vogt G., Woell S. and Argos P. (1997) Protein thermal stability, hydrogen bonds, and ion pairs. *J. Mol. Biol.* **269**: 631–643
- 18 Lazaridis T., Lee I. and Karplus M. (1997) Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. *Protein Sci.* **6**: 2589–2605
- 19 Usher K. C., De la Cruz A. F. A., Dahlquist F. W., Swanson R. V., Simon M. I. and Remington S. J. (1998) Crystal structures of CheY from *Thermotoga maritima* do not support conventional explanations for the structural basis of enhanced thermostability. *Protein Sci.* **7**: 403–412
- 20 Hough D. W. and Danson M. J. (1999) Extremoenzymes. *Curr. Opin. Chem. Biol.* **3**: 39–46
- 21 Kumar S., Tsai C. J. and Nussinov R. (2000) Factors enhancing protein thermostability. *Protein Eng.* **3**: 179–191
- 22 Szilagy A. and Zavodszky P. (2000) Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* **8**: 493–504
- 23 Salminen T., Teplyakov A., Kankare J., Cooperman B. S., Lahti R. and Goldman A. (1996) An unusual route to thermostability disclosed by the comparison of *Thermus thermophilus* and *Escherichia coli* inorganic pyrophosphatases. *Protein Sci.* **5**: 1014–1025
- 24 Karshikoff A. and Ladenstein R. (1998) Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. *Protein Eng.* **11**: 867–872
- 25 Haney P., Konisky J., Koretke K. K., Luthey-Schulten Z. and Wolyne P. G. (1997) Structural basis for thermostability and identification of potential active site residues for adenylate

- kinases from the archaeal genus *Methanococcus*. *Proteins* **28**: 117–130
- 26 Bogin O., Peretz M., Hacham Y., Korkhin Y., Frolow F., Kalb(Gilboa) A. J. et al. (1998) Enhanced thermal stability of *Clostridium beijerinckii* alcohol dehydrogenase after strategic substitution of amino acid residues with prolines from the homologous thermophilic *Thermanobacter brockii* alcohol dehydrogenase. *Protein Sci.* **7**: 1156–1163
- 27 Watanabe K., Hata Y., Kizaki H., Katsube Y. and Suzuki Y. (1997) The refined crystal structure of *Bacillus cereus* oligo-1,6-glucosidase at 2.0Å resolution: structural characterization of proline-substitution sites for protein thermostabilization. *J. Mol. Biol.* **269**: 142–153
- 28 Becktel W. and Schellman J. A. (1987) Protein stability curves. *Biopolymers* **26**: 1859–1877
- 29 Privalov P. L. (1990) Cold denaturation of proteins. *Crit. Rev. Biochem. Mol. Biol.* **25**: 281–305
- 30 Pace C. N. and Shaw K. L. (2000) Linear extrapolation method of analyzing solvent denaturation curves. *Proteins* **41**: 1–7
- 31 Cooper A. (1999) Thermodynamic analysis of biomolecular interactions. *Curr. Opin. Struct. Biol.* **3**: 557–563
- 32 Pace C. N., Hebert E. J., Shaw K. L., Schell D., Both V., Krajcikova D. et al. (1998) Conformational stability and thermodynamics of folding of ribonucleases Sa, Sa2 and Sa3. *J. Mol. Biol.* **279**: 271–286
- 33 Myers J. K., Pace C. N. and Scoltz J. M. (1995) Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**: 2138–2148
- 34 Edgcomb S. P. and Murphy K. P. (2000) Structural energetics of protein folding and binding. *Curr. Opin. Biotechnol.* **11**: 62–66
- 35 Ganesh C., Eswar N., Srivastava S., Ramakrishnan C. and Varadarajan R. (1999) Prediction of maximal stability temperature of monomeric globular proteins solely from amino acid sequence. *FEBS Lett.* **454**: 31–36
- 36 Beadle B. M., Baase W. A., Wilson D. B., Gilkes N. R. and Shoichet B. K. (1999) Comparing the thermodynamic stabilities of a related thermophilic and mesophilic enzyme. *Biochemistry* **38**: 2570–2576
- 37 Hollien J. and Marqusee S. (1999a) Structural distribution of stability in a thermophilic enzyme. *Proc. Natl. Acad. Sci. USA* **96**: 13674–13678
- 38 Hollien J. and Marqusee S. (1999b) A thermodynamic comparison of mesophilic and thermophilic ribonuclease H. *Biochemistry* **38**: 3831–3836
- 39 Li W.-T., Grayling R. A., Sandman K., Edmondson S., Shriver J. W. and Reeve J. N. (1998) Thermodynamic stability of archaeal histones. *Biochemistry* **37**: 10563–10572
- 40 Li W.-T., Shriver J. W. and Reeve J. N. (2000) Mutational analysis of differences in thermostability between histones from mesophilic and hyperthermophilic archaea. *J. Bacteriol.* **182**: 812–817
- 41 Wassenberg D., Welker C. and Jaenicke R. (1999) Thermodynamics of the unfolding of the cold-shock protein from *Thermotoga maritima*. *J. Mol. Biol.* **289**: 187–193
- 42 Petrosian S. A. and Makhatadze G. I. (2000) Contribution of proton linkage to the thermodynamic stability of the major cold-shock protein of *Escherichia coli* CspA. *Protein Sci.* **9**: 387–394
- 43 Pfeil W. (1998) Protein Stability and Folding: Collection of Thermodynamic Data, Springer Berlin
- 44 Frankenberg N., Welker C. and Jaenicke R. (1999) Does the elimination of ion pairs affect the thermal stability of cold shock protein from the hyperthermophilic bacterium *Thermotoga maritima*? *FEBS Lett.* **454**: 299–302
- 45 Mueller U., Perl D., Schmid F. X. and Heinemann U. (2000) Thermal stability and atomic-resolution crystal structure of the *Bacillus caldolyticus* cold shock protein. *J. Mol. Biol.* **297**: 975–988
- 46 Perl D., Mueller U., Heinemann U. and Schmid F. X. (2000) Two exposed amino acid residues confer thermostability on a cold shock protein. *Nat. Struct. Biol.* **7**: 380–383
- 47 Pace C. N. (2000) Single surface stabilizer. *Nat. Struct. Biol.* **7**: 345–346
- 48 Dahiyat B. I. (1999) In silico protein design. *Curr. Opin. Biotechnol.* **10**: 387–390
- 49 Gromiha M. M., Oobatake M. and Sarai A. (1999) Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins. *Biophys. Chem.* **82**: 51–67
- 50 Ladenstein R. and Antranikian G. (1998) Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. *Adv. Biochem. Eng. Biotechnol.* **61**: 37–85
- 51 Jaenicke R. and Bohm G. (1998) The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **8**: 738–748
- 52 Bernstein F. C., Koetzle T. F., Williams G. J., Meyer E. E. Jr., Brice M. D., Rogers J. R. et al. (1977) The Protein Data Bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* **112**: 535–542
- 53 Kelly C. A., Nishiyama M., Ohnishi Y., Beppu T. and Birktoft J. J. (1993) Determinants of protein thermostability observed in the 1.9-Å crystal structure of malate dehydrogenase from the thermophilic bacterium *Thermus flavus*. *Biochemistry* **32**: 3913–3922
- 54 Day M. W., Hsu B. T., Joshua-Tor L., Park J.-B., Zhou Z. H., Adams M. W. W. et al. (1992) X-ray crystal structures of the oxidized and reduced forms of the rubredoxin of the marine hyperthermophilic archaeobacterium *Pyrococcus furiosus*. *Protein Sci.* **1**: 1494–1507
- 55 Hiller R., Zhou Z. H., Adams M. W. W. and Englander S. W. (1997) Stability and dynamics in a hyperthermophilic protein with melting temperature close to 200°C. *Proc. Natl. Acad. Sci. USA* **94**: 11329–11332
- 56 Knegtel R. M. A., Wind R. D., Rozeboom H. J., Kalk K. H., Buitelaar R. M., Dijkhuizen L. et al. (1996) Crystal structure at 2.3Å resolution and revised nucleotide sequence of the thermostable cyclodextrin glycosyltransferase from *Thermobacterium thermosulfurigenes* EMI. *J. Mol. Biol.* **256**: 611–622
- 57 Kjeldgaard M., Nissen P., Thirup S. and Nyborg J. (1993) The crystal structure of elongation factor EF-TU from *Thermus aquaticus* in the GTP conformation. *Structure* **1**: 35–50
- 58 Yip K. S. P., Stillman T. J., Britton K. L., Artymiuk P. J., Baker P. J., Sedelnikova S. E. et al. (1995) The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure* **3**: 1147–1158
- 59 Adams M. W. W. (1993) Enzymes and proteins from organisms that grow near and above 100°C. *Annu. Rev. Microbiol.* **47**: 627–658
- 60 Klump H. H., Adams M. W. W. and Robb F. T. (1994) Life in the pressure cooker: the thermal unfolding of proteins from hyperthermophiles. *Pure Appl. Chem.* **66**: 485–489
- 61 Wigley D. B., Gamblin S. J., Turkenburg J. P., Dodson E. J., Piontek K., Muirhead H. et al. (1992) Structure of a ternary complex of an allosteric lactate dehydrogenase from *Bacillus stearothermophilus* at 2.5 Å resolution. *J. Mol. Biol.* **223**: 317–335
- 62 Holland D. R., Hausrath A. C., Juers D. and Matthews B. W. (1995) Structural analysis of zinc substitutions in the active site of thermolysin. *Protein Sci.* **4**: 1955–1965
- 63 Singleton P. and Sainsbury D. (eds) (1978) Dictionary of Microbiology and Molecular Biology, 2nd edn, Wiley, New York
- 64 Davies G. J., Gamblin S. J., Littlechild J. A. and Watson H. C. (1993) The structure of a thermally stable 3-phosphoglycerate

- kinase and a comparison with its mesophilic equivalent. *Proteins* **15**: 283–289
- 65 Auerbach G., Jacob U., Grottinger M., Schurig H. and Jaenicke R. (1997) Crystallographic analysis of phosphoglycerate kinase from the hyperthermophilic bacterium *Thermotoga maritima*. *Biol. Chem.* **378**: 327–329
  - 66 Jiang Y., Nock S., Nesper M., Sprinzl M. and Sigler P. B. (1996) Structure and importance of the dimerization domain in elongation factor Ts from *Thermus thermophilus*. *Biochemistry* **35**: 10269–10278
  - 67 Tsunasawa S., Izu Y., Miyagi M. and Kato I. (1997) Methionine aminopeptidase from the hyperthermophile archaeon *Pyrococcus furiosus*: molecular cloning and overexpression in *E. coli* of the gene and characteristics of the enzyme. *J. Biochem.* **122**: 843–850
  - 68 Gomes J., Gomes I., Kreiner W., Esterbauer H., Sinner M., and Steiner W. (1993) Production of high level of cellulase free and thermostable xylanase by a wild strain of *T. lanuginosus* using beechwood xylan. *J. Biotechnol.* **30**: 283–297
  - 69 Glaser P., Presecan E., Delepierre M., Surewicz W. K., Mantsch H. H., Barzu O. et al. (1992) Zinc, a novel structural element found in the family of bacterial adenylate kinases. *Biochemistry* **31**: 3038–3043
  - 70 Fukuyama K., Nagahara Y., Tsukihara T., Katsube Y., Hase T. and Matsubara H. (1988) Tertiary structure of *Bacillus thermoproteolyticus* [4Fe-4S] ferredoxin: evolutionary implications for bacterial ferredoxin. *J. Mol. Biol.* **199**: 183–193
  - 71 Obmolova G., Kuranova I. and Teplyakov A. (1993) Purification, crystallization and preliminary X-ray analysis of inorganic pyrophosphatase from *Thermus thermophilus*. *J. Mol. Biol.* **232**: 312–313
  - 72 Rypniewski W. R. and Evans P. R. (1989) Crystal structure of unliganded phosphofructokinase from *Escherichia coli*. *J. Mol. Biol.* **207**: 805–821
  - 73 Sako Y., Nomura N., Uchida A., Ishida Y., Morii Y., Koga Y., et al. (1996) *Aeropyrum pernix* gen. nov., sp. nov., a novel aerobic hyperthermophilic archaeon growing at temperatures up to 100 degrees C. *Int. J. Syst. Bacteriol.* **46**: 1070–1077
  - 74 Klenk H.-P., White O., Tomb J.-F., Clayton R. A., Nelson K. E., Ketchum K. A. et al. (1997) The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**: 364–370
  - 75 Smith D. R., Doucette-Stamm L. A., Deloughery C., Lee H., Dubois J., Aldredge T. et al. (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J. Bacteriol.* **179**: 7135–7155
  - 76 Bult C. J., White O., Olsen G.J., Zhou L., Fleischmann R. D., Sutton G. G. et al. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**: 1058–1073
  - 77 Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y., Yamamoto S. et al. (1998) Complete sequence and gene organization of the genome of a hyperthermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* **5**: 55–76
  - 78 Ruepp A., Graml W., Santos-Martinez M. L., Koretke K. K., Volker C., Mewes H. W. et al. (2000) The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407**: 508–513
  - 79 Deckert G., Warren P. V., Gaasterland T., Young W. G., Lenox A. L., Graham D. E. et al. (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392**: 353–358
  - 80 Nelson K. E., Clayton R. A., Gill S. R., Gwinn M. L., Dodson R. J., Haft D. H. et al. (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**: 323–329
  - 81 Kawashima T., Amano N., Koike H., Makino S. C., Higuchi S. Kawashima-Ohya Y. et al. (2000) Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*. *Proc. Natl. Acad. Sci. USA* **97**: 14257–14262
  - 82 Bernal A., Ear U. and Kyripides N. (2001) Genomes online database (GOLD): a monitor of genome projects world-wide. *Nucleic. Acids. Res.* **29**: 126–127
  - 83 Chakravarty S. and Varadarajan R. (2000) Elucidation of determinants of protein stability through genome sequence analysis. *FEBS Lett.* **470**: 65–69
  - 84 Thompson M. J. and Eisenberg D. (1999) Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. *J. Mol. Biol.* **290**: 595–604
  - 85 Zhang J. (2000) Protein-length distributions for the three domains of life. *Trends Genet.* **16**: 107–109
  - 86 Cambillau C. and Claverie J.-M. (2000) Structural and genomic correlates of hyperthermostability. *J. Biol. Chem.* **275**: 32383–32386
  - 87 Haney P. J., Badger J. H., Buldak G. L., Reich C. I., Woese C. R. and Olsen G. J. (1999) Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proc. Natl. Acad. Sci. USA* **96**: 3578–3583
  - 88 Kumar S., Ma B., Tsai C. J. and Nussinov R. (2000) Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers. *Proteins* **38**: 368–383
  - 89 Kumar S. and Nussinov R. (1999) Salt bridge stability in monomeric proteins. *J. Mol. Biol.* **293**: 1241–1255
  - 90 Kumar S. and Nussinov R. (2000) Fluctuations between stabilizing and destabilizing electrostatic contributions of ion pairs in conformers of the c-Myc-Max leucine zipper. *Proteins* **41**: 485–497
  - 91 Kumar S. and Nussinov R. (2001) Ion pairs and their stabilities fluctuate in NMR conformer ensembles of proteins. *Proteins* **43**: 433–454
  - 92 Xu D., Lin S. L. and Nussinov R. (1997) Protein binding versus protein folding: the role of hydrophilic bridges in protein associations. *J. Mol. Biol.* **265**: 68–84
  - 93 Horovitz A. and Fersht A. R. (1992) Co-operative interactions during protein folding. *J. Mol. Biol.* **224**: 733–740
  - 94 Marqusee S. and Sauer R. T. (1994) Contribution of a hydrogen bond/salt bridge network to the stability of secondary and tertiary structures in lambda repressor. *Protein Sci.* **3**: 2217–2225
  - 95 Pervushin K., Billeter M., Siegal G. and Wuthrich K. (1996) Structural role of a buried salt bridge in the 434 repressor DNA-binding domain. *J. Mol. Biol.* **264**: 1002–1012
  - 96 Lounnas V. and Wade R. C. (1997) Exceptionally stable salt bridges in cytochrome P450cam have functional roles. *Biochemistry* **36**: 5402–5417
  - 97 Barril X., Aleman C., Orozco M. and Luque F. J. (1998) Salt bridge interactions: stability of ionic and neutral complexes in the gas phase, in solution and in proteins. *Proteins* **32**: 67–79
  - 98 Maves S. A. and Sligar S. G. (2001) Understanding thermostability in cytochrome P450 by combinatorial mutagenesis. *Protein Sci.* **10**: 161–168
  - 99 Merz A., Knochel T., Jansonius J. N. and Kirschner K. (1999) The hyperthermostable indoleglycerol phosphate synthase from *Thermotoga maritima* is destabilized by mutational disruption of two solvent-exposed salt bridges. *J. Mol. Biol.* **288**: 753–763
  - 100 Strop P. and Mayo S. L. (2000) Contribution of surface salt bridges to protein stability. *Biochemistry* **39**: 1251–1255
  - 101 Kawamura S., Tanaka I. and Kimura M. (1997) Contribution of a salt bridge to the thermostability of DNA binding protein HU from *Bacillus stearothermophilus* determined by site directed mutagenesis. *J. Biochem.* **121**: 448–455
  - 102 Tanner J. J., Hecht R. M. and Krause K. L. (1996) Determinants of enzyme thermostability observed in the molecular

- structure of *Thermus aquaticus* D-glyceraldehyde-3-phosphate dehydrogenase at 2.5Å resolution. *Biochemistry* **35**: 2597–2609
- 103 Yip K. S. P., Britton K. L., Stillman T. J., Lebbink J., De Vos W. M., Robb F. T. et al. (1998) Insights into the molecular basis of thermal stability from the analysis of ion pair networks in the glutamate dehydrogenase family. *Eur. J. Biochem.* **255**: 336–346
- 104 Lebbink J. H. G., Knapp S., Oost J. van der, Rice D., Ladenstein R. and Vos W. M. de (1998) Engineering activity and stability of *Thermotoga maritima* glutamate dehydrogenase. I. Introduction of a six-residue ion-pair network in the hinge region. *J. Mol. Biol.* **280**: 287–296
- 105 Grimsley G. R., Shaw K. L., Fee L. R., Alston R. W., Huyghues-Despointes B. M., Thurlkill R. L. et al. (1999) Increasing protein stability by altering long-range coulombic interactions. *Protein Sci.* **8**: 1843–1849
- 106 Hendsch Z. S. and Tidor B. (1994) Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**: 211–226
- 107 Sheinermann F. B., Norel R. and Honig B. (2000) Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* **10**: 153–159
- 108 Bakker P. I. W. de, Hunenberger P. H. and McCammon J. A. (1999) Molecular dynamics simulations of the hyperthermophilic protein Sac7d from *Sulfolobus acidocaldarius*: contribution of salt bridges to thermostability. *J. Mol. Biol.* **285**: 1811–1830
- 109 Elcock A. H. and McCammon J. A. (1997) Continuum solvation model for studying protein hydration thermodynamics at high temperatures. *J. Phys. Chem. B* **101**: 9624–9634
- 110 Elcock A. H. (1998) The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J. Mol. Biol.* **284**: 489–502
- 111 Xiao, L. and Honig, B. (1999) Electrostatic contributions to the stability of hyperthermophilic proteins. *J. Mol. Biol.* **289**: 1435–1444
- 112 Sterner R. and Liebl W. (2001) Thermophilic adaptation of proteins. *Crit. Rev. Biochem. Mol. Biol.* **36**: 39–106
- 113 Kajander T., Kahn P. C., Passila S. H., Cohen D. C., Lehtio L., Adolfsen W. et al. (2000) Buried charged surface in proteins. *Structure* **8**: 1203–1214
- 114 Waldburger C. D., Jonsson T. and Sauer R. T. (1996) Barriers to protein folding: formation of buried polar interactions is a slow step in acquisition of structure. *Biochemistry* **93**: 2629–2634
- 115 Jaenicke R. (2000) Do ultrastable proteins from hyperthermophiles have high or low conformational rigidity? *Proc. Natl. Acad. Sci. USA* **97**: 2962–2964
- 116 Kumar S., Ma B., Tsai C. J., Sinha N. and Nussinov R. (2000) Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci.* **9**: 10–19
- 117 Tsai C. J., Kumar S., Ma B. and Nussinov R. (1999) Folding funnels, binding funnels and protein function. *Protein Sci.* **8**: 1179–1188
- 118 Ma B., Kumar S., Tsai C. J. and Nussinov R. (1999) Folding funnels and binding mechanisms. *Protein Eng.* **12**: 713–720
- 119 Villeret V., Clantin B., Tricot C., Legrain C., Roovers M., Stalton V. et al. (1998) The crystal structure of *Pyrococcus furiosus* ornithine carbamoyltransferase reveals a key role for oligomerization in enzyme stability at extremely high temperatures. *Proc. Natl. Acad. Sci. USA* **95**: 2801–2806
- 120 Vanhove M., Houba S., b1motte-Brasseur J. and Frere J. M. (1995) Probing the determinants of protein stability: comparison of class A  $\beta$ -lactamases. *Biochem. J.* **308**: 859–864.
- 121 Loladze V. V., Ibarra-olero B., Sanchez-Ruiz J. M. and Makhatadze G. I. (1999) Engineering a thermostable protein via optimization of charge-charge interactions on protein surface. *Biochemistry* **38**: 16419–16423
- 122 Spector S., Wang M., Carp S. A., Robblee J., Hendsch Z. S., Fairman R. et al. (2000) Rational modification of protein stability by mutation of charged surface residues. *Biochemistry* **39**: 872–879



To access this journal online:  
<http://www.birkhauser.ch>

---